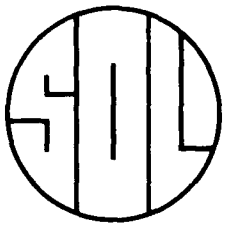
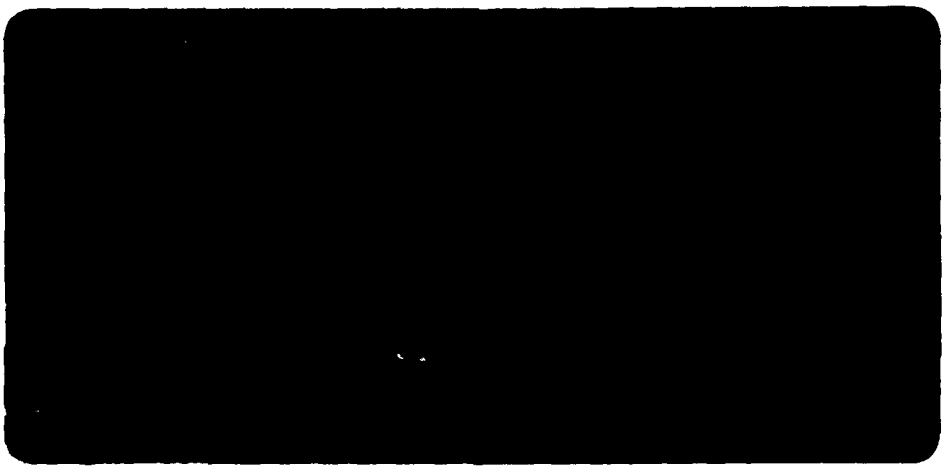
(4)



Systems
Optimization
Laboratory







Department of Operations Research
Stanford University
Stanford, CA 94305

4

SYSTEMS OPTIMIZATION LABORATORY
DEPARTMENT OF OPERATIONS RESEARCH
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305-4022

# Monte Carlo (Importance) Sampling within a Benders' Decomposition Algorithm for Stochastic Linear Programs

by
Gerd Infanger

TECHNICAL REPORT SOL 89-13

September 1989

Research and reproduction of this report were partially supported by the U.S. Department of Energy Grant DE-FG03-87ER25028; Office of Naval Research Contract N00014-89-J-1659; the National Science Foundation Grants DDM-8814253, DMS-8913089 and the Austrian Science Foundation, "Fonds zur Förderung der wissenschaftlichen Forschung," Grant J0323 - Phy.

Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author and do NOT necessarily reflect the views of the above sponsors.

# Monte Carlo (Importance) Sampling within a Benders' Decomposition Algorithm for Stochastic Linear Programs

Gerd Infanger

Stanford University

## Abstract

A method employing decomposition techniques and Monte Carlo sampling (importance sampling) to solve stochastic linear programs is described and applied to capacity expansion planning problems of electric utilities. We consider uncertain availability of generators and transmission lines and uncertain demand. Numerical results are presented.

Introduction

A stochastic linear program is a linear program whose parameters (coefficients, right hand sides) are uncertain. The uncertain parameters are assumed to be known only by their distributions. That means that the values of some functions are numerical characteristics of random phenomena, e.g. mathematical expectations of functions dependent on decision variables and random parameters (Kall. et. al. 1988).

Suppose a function $z = E\ C(V)$ is an expectation of a function $C(V^{\delta})$, $\delta \in \Omega$. V is a random parameter which has outcomes $V^{\delta}$. $\Omega$ is the set of all possible random events. It can be finite, infinite, discrete or continuous. In the continuous case the computation of the expected value requires to solve the integral:

$$E\ C(V) = \int C(V^{\delta})P(d\delta)$$

with P being the probability measure.

In a general case V would consist of several components, e.g. $V = (V_1,\ldots,V_h)$ with outcomes $V^{\delta}$ which we also will denote by lower case letters, e.g. $v = (v1,\ldots$ and $p(V^{\delta})$ alias p(v) would denote the corresponding density function. In this case the above mentioned integral takes the form of a multiple integral:

$$E\ C(V) = \int\ldots\int C(v)p(v)dv_1\ldots\ldots dv_h$$

In the case of $\Omega$ being discrete and finite the expectation can be computed by a multiple sum:

$$E\ C(V) = \sum_{v_1} \ldots \sum_{v_h} C(v)p(v)$$

The main difficulties in stochastic linear programming deal with the evaluation of the multiple integral or the multiple sum. The numerical computation of the expectation requires a large number of function evaluations and each function evaluation means a linear program to be solved. Different approaches attack this problem, e.g. Birge and Wallace (1988), Kall et. al. (1988), Rockafellar and Wets (1989) and others. We follow the concept of Dantzig et. al. (1989).

Two Stage Stochastic Linear Program

An important class of models concerns dynamic linear programs. Activities initiated at time t have coefficients at time t and t+1. Deterministic dynamic linear programs appear as staircase problems. The simplest staircase problem is that with two stages: X denotes the first, Y the second stage decision variables, A, b represent the coefficients and right hand sides of the first stage constraints and D, d concern the second period constraints together with B which couples the two periods. c, f are the objective function coefficients.

In the deterministic case c, f, A, b, B, D ,d are known with certainty to the planner. In the stochastic case, the parameters of the second stage are not known to the planner at time t=1, but will be known at time t=2. At time t=1 only the distribution of these parameters are assumed to be known. The second stage parameters can be seen as random variables which get certain outcomes with certain probabilities. We denote a certain outcome of these parameters with $\delta$ and the corresponding probability with $p(\delta)$, $\delta \in \Omega$ ,the set of possible outcomes.

$$
\begin{aligned}
\min\ z &= cX + E^{\delta}(fY^{\delta}) \\
\text{s/t}\quad & AX = b \qquad (1)\\
& B^{\delta}X + DY^{\delta} = d^{\delta} \\
& X, Y^{\delta} \geq 0, \quad \delta \in \Omega
\end{aligned}
$$

In (1) a two stage staircase problem is transformed into a two stage stochastic linear program and the parameters $d^{\delta}$ and $D^{\delta}$ being random variables. Given the two stage stochastic linear program one wants to make a decision X which is feasible for all scenarios and has the minimum expected costs.

We consider the case of $\Omega$ being discrete and finite, e.g. $\Omega = (1,\ldots,K)$, the parameter $\delta$ takes on K values. Then we can formulate an equivalent deterministic problem to the stochastic linear problem. This is tractable if K is small. For K=3 the deterministic equivalent problem is given in (2).

$$
\begin{aligned}
\min\ z &= cX + p^1fY^1 + p^2fY^2 + p^3fY^3 \\
\text{s/t}\quad & AX = b \qquad (2)\\
& B^1X + DY^1 = d^1 \\
& B^2X + DY^2 = d^2 \\
& B^3X + DY^3 = d^3 \\
& X, Y^1, Y^2, Y^3 \geq 0
\end{aligned}
$$

Two stage stochastic linear programs were first studied in Dantzig (1955) and then developed by many authors. The method which we want to apply here is using Benders (1962) decomposition. Van Slyke and Wets (1969) suggested to express the impact of the second period by a scalar $\theta$ and "cuts", which are necessary conditions to the problem and are expressed only in terms of the first period variables X and $\theta$. Benders decomposition splits the original problem into a master problem

and a subproblem which decomposes into a series of independent subproblems, one according to each $\delta \in \Omega$. The master problem, the sub problems and the cuts are represented in (3), (4) and (5).

The master problem:

$$\begin{aligned} \min\ & z_M = cX + \Theta \\ \text{s/t}\quad & AX = b \\ & G^l X + \alpha\Theta \geq g^l,\ l = 1,\ldots,L \\ & X,\ \Theta \geq 0 \end{aligned} \tag{3}$$

The sub problems:

$$\min\ z_S^{\delta} = p^{\delta} f Y^{\delta} \tag{4}$$

$$\text{s/t}\quad p^{\delta}\pi^{\delta}:\ DY^{\delta} = d^{\delta} + B^{\delta}X$$

$$Y^{\delta} \geq 0,\ \delta \in \Omega,\ \text{e.g. } \Omega = \{1,2,3\}$$

where $p^{\delta}\pi^{\delta}$ is the optimal dual solution of subproblem $\delta$.

The cuts:

$$g = \Sigma_{\delta}\ p^{\delta}\pi^{\delta}d^{\delta} = E(\pi^{\delta}d^{\delta}) \tag{5}$$

$$G = \Sigma_{\delta}\ p^{\delta}\pi^{\delta}B^{\delta} = E(\pi^{\delta}B^{\delta})$$

$\alpha = 0$ ... feasibility cut
$\alpha = 1$ ... optimality cut

Solving the master problem we obtain a solution X. Given X we can solve K subproblems $\delta \in \Omega$ to compute a cut. The cut is a lower bound on the expected value of the second stage costs represented as a function of X. Cuts are sequentially added to

the master problem and new values of X are obtained until the optimality criterion is met. We distinguish between two types of cuts, feasibility cuts and optimality cuts. The first refers to infeasible subproblems for a given X and the latter to feasible and optimum subproblems, given X.

If the expected values z, G, and g are computed exactly, that is, by evaluating all scenarios $\delta \in \Omega$, we refer to it as the universe case. As we will see later the number of scenarios easily gets out of hand and it is not always possible to solve the universe case. Therefore methods are sought which guarantee a satisfying solution without solving the universe case. Employing Monte Carlo methods seems to be a promising approach.

Monte Carlo Sampling

Each iteration of Benders' decomposition requires the computation of expected values ,e.g. the subproblem costs, the coefficients and right hand sides of the cuts. For each outcome $\delta \in \Omega$ a linear program has to be solved. The expected value of e.g. the subproblem costs is denoted by

$$z_S = E\ C(V^\delta) = E\ fY^\delta \quad , \delta \in \Omega.$$

The number of elements of $\Omega$ is determined by the dimensionality of the stochastic vector $V = (V_1,\ldots,V_h)$. Typically the dimension h of V is quite large. E.g. in expansion planning problems of electric power systems a component of V denotes the availability of a type of generators or a demand of power in a node of a multi area supply network or the availability of a type of transmission line connecting two nodes. Consider several nodes and arcs and one demand and some options of generators in each node. The number of scenarios K in the universe case gets quickly out of hand, even if the distribution of each component of V is determined by just a small number $K^i$ of

discrete points. Suppose e.g. $h = 20$ and $K^i = 5$. Then the total number of terms in the expected value calculations is $K = 5^{20} \doteq 10^{12}$, which is not practically solvable by direct summation (Glynn in Dantzig et al. (1989)). Monte Carlo methods are recommended to compute multiple integrals or multiple sums for h large (Davis and Rabinowitz (1984)). See Hammersly and Handscomb (1964) for a description of Monte Carlo sampling techniques.

Crude Monte Carlo

Suppose $V^\delta$, $\delta = 1,\ldots,n$ are scenarios, sampled independently from their joint probability mass function, then $C^\delta = C(V^\delta)$ are independent random variates with expectation z. (The subscript S is suppressed now as there is no danger of confusion.)

$$\acute{z} = (1/n) \sum_{\delta=1}^{n} C^\delta \qquad (6)$$

is an unbiased estimator of z and its variance

$$\sigma^2_{\acute{z}} = \sigma^2/n$$
$$\sigma^2 = \mathrm{var}(C(V)).$$

Thus the standard error is decreasing with sample size n by $n^{-0.5}$. *The convergence rate of $\acute{z}$ to z is independent of the* dimension h of the random vector V.

Importance Sampling

We rewrite

$$z = \Sigma_V \, c(v)p(v) \quad = \quad \Sigma_V \, c(v)p(v)q(v)/q(v)$$

by introducing a probability mass function q(v). We can see q as a probability mass function of a random vector W, therefore

$$z = E \; C(W)p(W)/q(W)$$

and we obtain a new estimator

$$z = (1/n) \sum_{\delta=1}^{n} C(W^{\delta})p(W^{\delta})/q(W^{\delta})$$

which has a variance of

$$var(z) = (1/n) \; \Sigma_W (C(w)p(w)/q(w) - z)^2 \; q(w)$$

Choosing $q^*(w) = C(w)p(w)/(\Sigma_W \, C(w)p(w))$ would lead to var(z) = 0, that means we could get a perfect estimate of the multiple sum just by one single observation. However this is practically useless, since to sample C.p/q we have to know q and to determine q we need to know $z = \Sigma_W \, c(w)p(w)$, which we eventually want to compute. Nevertheless this result helps to derive some heuristics of how to choose q: It should be approximately proportional to the product C(w)p(w) and have a form which can be integrated

theoretically. For the theory of importance sampling we refer to Glynn and Iglehart (1988) and Dantzig and Glynn (1989a).

P.Glynn and M.Nakayama in Dantzig et. al. (1989) developed an importance sampling scheme using an additive model to approximate the cost function E C(V):

$$C(v) = \sum_{i=1}^{h} C_i(v_i).$$

Actually C(v) is achieved by a marginal cost model, considering marginal costs in each dimension i of V.

$$C(v) = C(\tau) + \sum_{i=1}^{h} M_i(v_i) \qquad (7)$$

$$M_i(v_i) = C(\tau_1,..,\tau_{i-1},v_i,\tau_{i+1},..,\tau_d) - C(\tau)$$

$\tau = (\tau_1,...,\tau_h)$ can be any arbitrary chosen point out of the set of values $v_i$, i = 1,..,h. For example we choose $\tau_i$ that outcome of $V_i$ which leads to the respectively lowest costs. In the context of expansion planning of power systems this means selecting respectively lowest demands and highest availabilities of generators and transmission lines.

Defining

$$\bar{M}_i = E\ M_i(V_i) \qquad (8)$$

and

$$F(v) = ((C(v) - C(\tau))/ \sum_{i=1}^{h} M_i(v_i) \qquad (9)$$

we can express the expected value of the costs in the following form, e.g. in the case of h = 3:

$$\begin{aligned} z = C(\tau) &+ \bar{M}_1 \Sigma_v F(v) \, p_1(v_1)M_1(v_1)/\bar{M}_1)p_2(v_2)p_3(v_3) \\ &+ \bar{M}_2 \Sigma_v F(v) \, p_1(v_1)(p_2(v_2)M_2(v_2)/\bar{M}_2)p_3(v_3) \\ &+ \bar{M}_3 \Sigma_v F(v) \, p_1(v_1)p_2(v_2)(p_3(v_3)M_3(v_3)/\bar{M}_3) \end{aligned} \qquad (10)$$

Note that this formulation consists of a constant term and h expectations. Given a fixed sample size n we partition n into $n^i$, i = 1,..h sub-samples , such that $\Sigma\, n_i = n$ and $n_i \geq 1$, i = 1,...,n and $n_i$ being approximately proportional to $M_i$. The h expectations are separately approximated by sampling using marginal densities. The i-th expectation corresponds of course to the i-th component of V. Generating sample points in the i-th expectation we use the importance density $(p_i M_i/\bar{M}_i)$ for sampling the i-th component of V and the original densities for any other components. Denoting

$$\mu_i = (1/n_i) \sum_{j=1}^{n_i} F(v_j) \qquad (11)$$

the estimate of the i-th sum, we obtain

$$\bar{z} = C(\tau) + \sum_{i=1}^{h} \bar{M}_i \mu_i \ , \qquad (12)$$

the expected value of the second stage costs C(V).

Let $\sigma_i^2$ be the sample variance of the i-th expectation,where $\sigma_i^2 = 0$ if $n_i = 1$. The variance of the mean is then given by

$$\sigma_{\bar{z}}^2 = \sum_{i=1}^{h} \bar{M}_i^2 \sigma_i^2 \ / \ n_i \qquad (13)$$

Using importance sampling one can achieve significant variance reduction. The experiment of M. Nakayama in Dantzig et al. (1989) claims a variance reduction of 1:20000 using importance sampling versus crude Monte Carlo sampling: For a given and optimal X the second stage costs of a multi area expansion planning model with 192 universe scenarios were sampled with a sample size of 10 using both methods and the results compared.

The derivation above concerned the expected second stage costs z. To derive a cut we use the same framework analogously. Note that a cut is defined as an outer linearization of the second stage costs represented as a function of X, the first stage variables. At X, the value of the cut is exactly the expected second stage costs z. Therefore we can employ directly the cost approximation scheme and the importance distribution to compute the parameters of a cut. We define

$$F^G(v) = ((\pi B)(v) - (\pi B)(\tau))/ \sum_{i=1}^{h} M_i(v_i) \qquad (14)$$

$$F^g(v) = ((\pi d)(v) - (\pi d)(\tau))/ \sum_{i=1}^{h} M_i(v_i) \qquad (15)$$

and get e.g. in the case of h = 3

$$\begin{aligned} G = (\pi B)(\tau) &+ \bar{M}_1 \Sigma_v F^G(v) \; (p_1(v_1)M_1(v_1)/\bar{M}_1)p_2(v_2)p_3(v_3) \\ &+ \bar{M}_2 \Sigma_v F^G(v) \; p_1(v_1)(p_2(v_2)M_2(v_2)/\bar{M}_2)p_3(v_3) \\ &+ \bar{M}_3 \Sigma_v F^G(v) \; p_1(v_1)p_2(v_2)(p_3(v_3)M_3(v_3)/\bar{M}_3) \end{aligned} \qquad (16)$$

$$\begin{aligned} g = (\pi d)(\tau) &+ \bar{M}_1 \Sigma_v F^g(v) \; (p_1(v_1)M_1(v_1)/\bar{M}_1)p_2(v_2)p_3(v_3) \\ &+ \bar{M}_2 \Sigma_v F^g(v) \; p_1(v_1)(p_2(v_2)M_2(v_2)/\bar{M}_2)p_3(v_3) \\ &+ \bar{M}_3 \Sigma_v F^g(v) \; p_1(v_1)p_2(v_2)(p_3(v_3)M_3(v_3)/\bar{M}_3) \end{aligned} \qquad (17)$$

the coefficients and the right hand side of a cut. We compute the expected values again by sampling using the same sample points as at hand from the computation of z.

Using Monte Carlo sampling we obtain $\tilde{z}$ $(\sigma_{\tilde{z}})$, $\tilde{G}$, $\tilde{g}$, which are approximations of the expected values z = E c(V), G = E $(\pi B)(V)$, g = E $(\pi d)(V)$. The impact of using approximations instead of the exact parameters on the Benders decomposition algorithm is analyzed in the following section.

Benders' Decomposition

In the following we will derive the main steps of Benders decomposition algorithm for two stage stochastic linear programs considering the "universe" case, which gives the exact solution of the equivalent deterministic problem ("certainty equivalent"). We will then analyze the impact of sampling of subproblems on Benders decomposition. See Geoffrion (1970) for a derivation of Benders decomposition algorithm.

Given the equivalent deterministic problem in (2) and assuming that K = 3 describes the universe case, we rewrite the problem applying projection onto the X variables and obtain (18). We assume for simplicity that (2) is feasible and has a finite optimum solution.

$$
\begin{array}{l}
\min z = \\
cX + \mathrm{Inf}\{p^1 f Y^1 + p^2 f Y^2 + p^3 f Y^3\} \\
AX = b \qquad DY^1 = d^1 + B^1X \\
X \geq 0 \qquad DY^2 = d^2 + B^2X \\
\qquad\qquad DY^3 = d^3 + B^3X \\
\qquad\qquad Y^1, Y^2, Y^3 \geq 0
\end{array}
\tag{18}
$$

The infimal value function in (18) corresponds to the following primal linear problem (19):

$$
\begin{array}{lll}
\min z_p = p^1 f Y^1 + p^2 f Y^2 + p^3 f Y^3 = E^{\delta}(f Y^{\delta}) & & (19) \\
p^1\pi^1: \quad DY^1 = d^1 + B^1X & & \\
p^2\pi^2: \quad DY^2 = d^2 + B^2X & & \\
p^3\pi^3: \quad DY^3 = d^3 + B^3X & & \\
Y^1, Y^2, Y^3 \geq 0 & &
\end{array}
$$

and to the dual linear problem (20):

$$\max z_D = \qquad (20)$$
$$p^1\pi^1(d^1+B^1X) + p^2\pi^2(d^2+B^2X) + p^3\pi^3(d^3+B^3X)$$

$$\begin{aligned} \pi^1 D &= f \\ \pi^2 D &= f \\ \pi^3 D &= f \\ \pi^1,\ \pi^2,\ \pi^3 &\geq 0 \end{aligned}$$

The primal problem is parameterized in the right hand side by X. The assumption (2) being finite implies that (19) is finite for at least one value of X. Applying the Duality Theorem of Linear Programming we state that (20) has to be feasible. The feasibility conditions

$$\pi^\delta D - f = 0$$

indicate that the feasible region $\{\pi^\delta | \pi^\delta D - f \leq 0\}$ is independent of X and just repeated for each scenario $\delta \in \Omega$.

The assumption (2) being feasible requires feasibility of the primal problem (19) for at least one X. By the Duality Theorem again (20) has to be finite. Let $\pi^j$, j=1,...,p be the extreme points and $\pi^j$, j=p+1,...,p+q be representatives of the extreme rays of the feasible region of (9). Problem (20) is finite if and only if

$$\pi_j(d^\delta + B^\delta X) \leq 0, \quad j = p+1,\ldots,p+q \qquad (21)$$
$$\delta \in \Omega$$

We append constraints (21) to problem (18) to ensure that the problem is bounded.

We call

$$z_D{}^{\delta *} = \max_{1 \le j \le p} \; p^{\delta} \pi_j{}^{\delta} (d^{\delta} + B^{\delta} X) = z_D{}^{\delta} (\pi^{\delta *}) \qquad (22)$$

the optimum second stage costs for given $\hat{X}$ in scenario $\delta$. $\pi^{\delta *}$ denotes the optimum dual variables for scenario $\delta$ selected from the set $\pi_j{}^{\delta}$, j=1,..,p. It is clear that

$$z_D{}^{*} = \sum_{\delta \in \Omega} z_D{}^{\delta *} \qquad (23)$$

Let $\pi^{1*}, \pi^{2*}, \pi^{3*}$ be the vector of optimum dual variables of the second stage problem, given $\hat{X}$, we outer linearize the infimal value function in (18). By the dual problem (19) we obtain:

$$z^{D*} = p^1 \pi^{1*} (d^1 + B^1 X) + p^2 \pi^{2*} (d^2 + B^2 X) + p^3 \pi^{3*} (d^3 + B^3 X) \qquad (24)$$

and

$$z^D (\pi^{1*}, \pi^{2*}, \pi^{3*}, \hat{X}) \ge z^D (\pi^{1*}, \pi^{2*}, \pi^{3*}, X) \qquad (25)$$

formulates the main property of the outer linearization. This property enables us to rewrite problem (18) by expressing the infimal value function by the outer linearized dual problem.

$$\begin{aligned} \min z = \; & cX \qquad + z^D(X) \\ & AX = b \\ & X \geq 0 \end{aligned} \tag{26}$$

$$z^D(X) =$$

$$\max_{1 \leq j \leq p,} \quad p^1\pi_j{}^1(d^1+B^1X) + p^2\pi_j{}^2(d^2+B^2X) + p^3\pi_j{}^3(d^3+B^3X)$$

Using $\Theta$ as the greatest lower bound the problem can be represented in the following form:

$$\begin{aligned} \min z = \; & cX + \Theta \\ & AX \quad = b \\ & X \quad \geq 0 \end{aligned} \tag{27}$$

$$\Theta \geq p^1\pi_j{}^1(d^1+B^1X) + p^2\pi_j{}^2(d^2+B^2X) + p^3\pi_j{}^3(d^3+B^3X)$$

$$j = 1,\ldots,p$$

$$\pi_j(d^\delta + B^\delta X) \leq 0, \quad j = p+1,\ldots,p+q$$

$$\delta \in \Omega$$

Relaxation is applied to solve problem (12).

To test a solution $(X, \Theta)$ one solves problem (19) or problem (20), actually by solving

$$z_S{}^{\delta *} = \min z_P{}^{\delta} = fY^{\delta} \tag{28}$$

$$\text{s/t} \quad \pi^{\delta}: \quad DY^{\delta} = d^{\delta} + B^{\delta}X$$

$$Y^{\delta} \geq 0, \qquad \delta \in \Omega$$

or by solving

$$z_S{}^{\delta *} = \max z_D{}^{\delta} = \pi^{\delta}(d^{\delta}+B^{\delta}X) \tag{29}$$

$$\pi^{\delta}D = f$$

$$\pi^{\delta} \geq 0, \qquad \delta \in \Omega$$

If primal infeasibility or dual unboundness is detected, a feasibility cut

$$\pi^{\delta}(d^{\delta} + B^{\delta}X) \leq 0 \tag{30}$$

is added to the master problem. If all primal problems are feasible or all dual problems unbounded an optimality cut

$$\theta \geq \sum_{\delta \in \Omega} p^{\delta}\pi^{\delta}(d^{\delta}+B^{\delta}X) \tag{31}$$

is added to the master problem.

In the l-th iteration

$$\underline{z}^{l} = z_M{}^{l*} = cX^{l*} + \theta^{l*} \tag{32}$$

is defined to be a lower bound and

$$\bar{z}^l = \min\{z^{l-1},\ cX^{l*} + z_S{}^*\},\ z^0 = \infty \qquad (33)$$

to be an upper bound to the solution of the problem. If

$$(\bar{z} - \underline{z})/\underline{z} \leq TOL,$$

where TOL is a given tolerance, the problem is said to be solved with a sufficient accuracy.

Probabilistic Cuts

Employing Monte Carlo sampling techniques means not to solve all problems $\delta \in \Omega$, but solving problems $\delta \in S$, S being a subset of $\Omega$. Instead of the exact expected values $z_S$, G, g we compute the estimates $\tilde{z}_S$, $\tilde{G}$, $\tilde{g}$ by weighted sums. Suppose e.g. in problem (2) S is the set of problems {1,2} out of $\Omega$ = {1,2,3}. For example in the case of crude Monte Carlo sampling scenarios $\delta \in S$ are sampled according to the probability mass function $p^\delta$ and an approximation of the expected value is obtained by computing the mean of the samples, e.g. $z_S = (z^1 + z^2)/2$. Referring to (2), a cut obtained by crude Monte Carlo sampling would be computed as

$$\theta \geq (p^1/(p^1+p^2))(d^1+B^1X) + (p^2/(p^1+p^2))(d^2+B^2X).$$

Suppose $\tilde{\theta}$ is the approximation of $\theta$, the exact outer linearization of the second stage costs. The difference $\theta - \tilde{\theta}$ describes the error of the approximation. Comparing the primal (19) and the dual (20) problem for $\Omega$ and S we see

$$\theta - \tilde{\theta}(p^1 + p^2) = z_p{}^{3*}$$

The error $(\Theta - \Theta) = z_p{}^{3*} - p^3\Theta$ is constant with respect to X and affects therefore just the right hand side $g^1$ of a cut. See also *Dantzig and Glynn (1989b)* in this respect. In general S is a sufficient large subset of $\Omega$, $\Theta$ is computed by sampling methods and the error is derived from the variance of the sample mean and assumed to be small.

Cuts computed by samples do not necessarily meet the condition of outer linearization (25). Violating this condition a cut intersects and separates parts of the feasible region of the second stage problem. A sampled cut is therefore not a valid cut. The right hand sides of cuts obtained by sampling can be seen as stochastic parameters. We assume normal distributions defined by the means $\hat{g}^1$ and the standard deviations $\sigma_{\tilde{g}}{}^1$. We know that $\sigma_{\tilde{g}} = \sigma_{\tilde{z}}$, the standard deviation of the second stage costs.

Upper and Lower Bounds

For random right hand sides $g^1$ in Benders' master also the upper and lower bounds of the problem are probabilistic. The standard deviation of the upper bound, $\sigma_{\bar{z}}{}^1$ is given by $\sigma_{\bar{z}}{}^1 = \sigma_{zS1}$ the standard deviation of the subproblems costs.

To determine the standard deviation of the lower bound consider the master problem at iteration L:

$$
\begin{aligned}
&\min z_M = cX + \Theta && && (34)\\
&\text{s/t } \pi^0: \quad AX && = b\\
&\qquad \pi^1: \quad G^1X + \alpha\Theta && = g^1 \ (\sigma_{\tilde{z}}{}^1)\\
&\qquad \pi^L: \quad G^LX + \alpha\Theta && = g^L \ (\sigma_{\tilde{z}}{}^L)\\
&\qquad\qquad X, \Theta && \geq 0 \ ,
\end{aligned}
$$

where L cuts have been added to the originally relaxed master problem. The right hand sides are independent stochastic parameters, assumed normally distributed. We assume independence as the cuts are generated from independent samples, neglecting the dependency that $X_l$, $l = 1,..,L$ are weakly connected by the Benders' algorithm. Under these assumptions we experimentally obtain a distribution of $z_M^*$ by drawing N samples $j = 1,..,N$ from the normal distributions of $g^l$. Varying the right hand sides $(g^1,..,g^L)_j$, according to the samples $j = 1,..,N$ and solving the master problem for each $j = 1,..,N$ we obtain solutions $z_M^*{}_j$, $j = 1,..,N$, which determine the distribution of the lower bound. Assuming normal distribution, we compute the variance

$$\sigma^2_{zM} = \text{var}(z_M^*) = (1/N-1) \sum_{j=1}^{N} (z_M^*{}_j - z_M^*)^2 \qquad (35)$$

Solving the master problem N times to obtain an estimate of the lower bound variance is very expensive. We assume that the standard deviations $\sigma_z{}^l$, $l=1,..,L$ are small and all solutions $z_M^*j$, $j = 1,..,N$ have the same basis. Then we can compute the sample points from the dual objective function:

$$\Delta_j = \sum_{l=1}^{L} \pi^l (g^l{}_j - g^l) \qquad (36)$$

$$\sigma^2_{zM} = (1/N-1) \sum_{j=1}^{N} (\Delta_j)^2 \qquad (37)$$

or directly:

$$\sigma^2_{zM} = \sum_{l=1}^{L} \pi_l^2 \, \sigma_z^{L2} \qquad (38)$$

Stopping Rule

The accuracy of the solution is influenced by the accuracy of the cuts determined by the variances $\sigma^2_{\tilde{z}^l}$ and is a function of the sample size n. Given a fixed sample size n, the Benders algorithm

has to proceed until the maximum accuracy is achieved. Operating with probabilistic bounds the maximum accuracy is reached if the upper and lower bounds are identical in distribution. See Dantzig and Glynn (1989b) for an appropriate rule to determine this property. At least no improvement is possible anymore if the upper and the lower bound intersect. Employing this rule we thus define

$$\bar{z} - \underline{z} \leq 0 \qquad (39)$$

as the stopping criteria.

The Accuracy of the Solution

The accuracy of the solution can be estimated from the distribution of the lower bound of the problem after the last iteration. Given $z_M$ and $\sigma_{zM}$, in the last iteration, we compute a confidence interval e.g.

$$z_{c95} = z_{0.95}\ \sigma_z^M,$$

$z_{0.95}$ being the 95% quantile of the standard normal distribution.

If

$$z_{c95}/z_M^* \leq TOL, \tag{40}$$

the obtained solution is satisfying. The quality criterion TOL is understood here as a confidence interval. Otherwise the sample size has to be increased and the problem has to be solved again with the increased sample size.

Improvement of the Solution

Suppose the solution with a certain sample size was not satisfying. Instead of starting from the beginning with an increased sample size we want to use the information, which we have already collected. To do this, we look for the binding cuts in the final solution, increase the sample size and recompute the binding cuts at the same $\hat{X}^l$, they were originally computed. The enlarged sample size leads to smaller variances $\sigma^2_{\tilde{z}^l}$ of the binding cuts and eventually to a smaller confidence interval of the final solution. Solving the master problem again with the improved binding cuts will in general not result in an intersection of the lower and upper bound. Therefore some more iterations are necessary to obtain the optimal solution according to the increased sample size. This improvement procedure is employed iteratively until a satisfying solution is obtained. We can state now the algorithm as follows:

The Algorithm

Step 0 Initialize
$l = 0$, $z^0 = \infty$

Step 1 Solve the relaxed master problem and obtain a lower bond: $\underline{z}^l = cX^l + \theta^l$

Step 2 $l = l + 1$
Solve subproblems and obtain an upper bound: $\bar{z}_l = \min\{\bar{z}_{l-1}, cX_l + z_S^*\}$, compute and add a cut to the master problem, using Monte Carlo (importance) sampling.

Step 3 Solve the master problem and obtain a lower bond: $\underline{z}^l = cX^l + \theta^l$

Step 4 If $(\bar{z} - \underline{z} \leq 0)$ go to step 2

Step 5 Compute confidence interval $z_{c95}$ and obtain a solution: $z$, $X$, $\theta$

Step 6 If $(z_{c95}/z_M^* \leq TOL)$ stop, otherwise go to step 7

Step 7 Increase sample size and initialize $z^0 = \infty$

Step 8 Recompute binding cuts
Upper bound: $\bar{z}_l = \min\{\bar{z}_{l-1}, cX_l + z_S^*\}$,

Step 9 Go to step 3

Numerical Results

The method has been implemented in an APL environment and tested on small problems:

The test problem consists of a capacity expansion planning problem of electric utilities. There are two types of generators with different investment and operations costs, which can be built and operated in a way to meet the demand, given by a load duration curve of three load levels: base, medium and peak load. Both the availability of power from the generators and the loads of the three load levels are considered to be uncertain. We assume a discrete distribution of four outcomes of the first, five outcomes of the second generator and four outcomes of each of the demand. The model is formulated as a complete recourse model, that means we ensure feasibility of the subproblems for any X. "Unserved demand" can be purchased with costs, higher than the costs of production (penalty costs). The generators have intersecting costs functions so that building of both generators is reasonable. The building of the generators is in competition with the purchase of load.

The assumptions on the stochastic variables imply 1280 possible outcomes, i.e. 1280 subproblems have to be solved in each iteration of the Benders decomposition in the universe case. We compare the universe solution (the test problem is small enough to solve the universe problem) with solutions gained by the importance sampling algorithm.

Table 1 shows the results in the case of 20 samples out of the possible 1280 combinations and without an improvement phase. 100 replications of the same experiment were run to get statistical information about the accuracy of the solution and the estimate of the accuracy of the solution.

The mean of the objective function value (total costs) differs from the universe solution by 0.2%. From the distribution of the optimum objective function value given by the 100 replications a 95% confidence interval can be computed: plus

minus 1.5%. A 95% confidence interval of each solution of the 100 replications is estimated. The mean of all confidence intervals is 1.4%, which is a slight underestimation of the true confidence interval.This error is caused by the estimation method. The estimation method however improves in accuracy with decreasing confidence intervals of the solution. The coverage rate of 92% expresses that in 92% of the 100 replications the correct answer of the universe solution is covered by the estimated confidence interval. This again shows that we are slightly underestimating: if the computation the 95% confidence interval was exact, we would expect a coverage rate of 95%.

The bias and the confidence interval of the optimum strategies (the loads X to be installed) are larger than those of the optimum objective function value. The optimum seems to be flat: several different strategies lead close to the optimum costs. Confidence intervals of 52% and 48% are computed.

In the above example a sample size of 20 samples was chosen. Additional computational effort is also needed to obtain the importance distribution, e.g. 17 subproblems have to be solved in each iteration to obtain the marginal costs $M_i$. Compared to the universe solution the method e.g. achieves with 2.9% of computational effort a solution which is with 95% confidence within an interval of plus minus 1.5% of the correct answer.

Table 2 represents computational results for an enlarged model: We consider a triangular network of tree nodes connected by transmission lines. There is a load in each node and one type of generator in two of the nodes. The generators in the two nodes and the transmission lines have to be built to serve the demand. The generators and the transmission lines respectively differ in investment and operation costs. We consider two time periods (in the sense of "Here and Now" decision making, and independent in the uncertain parameters), dependent in the deterministic part of the problem. Thus there are 10 decision variables in the first stage (master) problem. The distributions of the uncertain parameters are identical to the first test problem. The problem

is solved with 20 and 40 samples out of 1280 universe scenarios using importance sampling and the results are compared. The statistical information is obtained by respectively 30 replications of the same experiment.

One can see decreasing bias, decreasing confidence intervals and improving estimations of the confidence intervals with increased sample size. A larger number of binding cuts (e.g. depending on the number of first stage variables and thus the problem size) in the final solution implies more error. Therefore a higher sample size is required to obtain the same accuracy of the solution, compared to smaller problems.

Table 1: 20 samples

| | correct | mean | 95%conf % | bias % |
|---|---|---|---|---|
| #iter | | 7.9 | | |
| G1 | 1800.0 | 1568.8 | 52.0 | -12.8 |
| G2 | 1571.4 | 1793.9 | 47.7 | 14.2 |
| theta | 13513.7 | 13922.8 | 15.9 | 3.0 |
| obj | 24642.3 | 24682.7 | 1.5 | 0.2 |
| est. conf | | 1.4 | | |
| coverage | | 0.92 | | |

Table 2: 20 samples

| | | correct | mean | 95%conf % | bias % |
|---|---|---|---|---|---|
| | #iter | 17 | 13.8 | | |
| period 1 | G1 | 2559 | 1656 | 116.0 | -35.3 |
| | G2 | 5082 | 5577 | 26.7 | 9.7 |
| | T1 | 1369 | 1860 | 65.9 | 35.8 |
| | T2 | 1119 | 1331 | 140.2 | 19.0 |
| | T3 | 0 | 72 | 435.2 | |
| | theta1 | 13336 | 16072 | 29.4 | 20.5 |
| period2 | G1 | 1889 | 2352 | 149.6 | 24.5 |
| | G2 | 7624 | 7759 | 43.3 | 1.8 |
| | T1 | 2787 | 2307 | 77.9 | -17.2 |
| | T2 | 1801 | 1721 | 71.0 | -4.4 |
| | T3 | 0 | 212 | 386.5 | |
| | theta2 | 28058 | 26712 | 29.0 | -4.8 |
| | obj | 99325 | 101199 | 3.7 | 1.9 |
| | est. conf | | 2.4 | | |
| | coverage | | 0.6 | | |

Table 2b: 40 samples

| | | correct | mean | 95%conf % | bias % |
|---|---|---|---|---|---|
| | #iter | 17 | 14.7 | | |
| period 1 | G1 | 2559 | 1189 | 58.0 | -53.5 |
| | G2 | 5082 | 5810 | 17.3 | 15.7 |
| | T1 | 1369 | 1744 | 55.6 | 27.4 |
| | T2 | 1119 | 1290 | 49.1 | 15.2 |
| | T3 | 0 | 213 | 310.2 | |
| | theta1 | 13336 | 17123 | 10.7 | 28.4 |
| period2 | G1 | 1889 | 1445 | 102.5 | -23.5 |
| | G2 | 7624 | 8670 | 17.8 | 13.7 |
| | T1 | 2787 | 3102 | 44.7 | 11.3 |
| | T2 | 1801 | 1758 | 50.8 | -2.4 |
| | T3 | 0 | 226 | 395.0 | |
| | theta2 | 28058 | 27196 | 16.8 | -3.1 |
| | obj | 99325 | 100347 | 2.3 | 1.0 |
| | est. conf | | 2.1 | | |
| | coverage | | 0.7 | | |

Acknowledgements

The research leading to this paper was conducted while the author was a visiting scholar at the Department of Operations Research at Stanford University. The author wants to thank Professor George B. Dantzig for his professional and personal support during this visit. The author also wants to thank Peter Glynn and John Stone for valuable discussions on this topic and Mordecai Avriel for helpful suggestions concerning this paper.

Literature

Benders,J.F.(1962):
Partitioning Procedures for Solving Mixed- Variable Programming Problems,Numerische Mathematic 4, 238-252.

Birge,J.R. and Wallace,S.W. (1988):
A Separable Piecewise Linear Upper Bound for Stochastic Linear Programs, SIAM J. Control and Optimization, 26, 5.

Dantzig G.B.(1955):
Linear Programming under Uncertainty, Management Science 1, 197-206

Dantzig G.B. and Glynn P.W. (1989a):
Decomposition and Parallel Processing Techniques for Large Scale Electric Power System Planning Under Uncertainty, paper presented at Workshop on Resource Planning Under Uncertainty, Stanford University.

Dantzig G.B. and Glynn P.W. (1989b):
Parallel Processors for Planning Under Uncertainty, Technical Report SOL 88-8R, Department of Operations Research, Stanford University.

Dantzig,G.B., Glynn,P.W., Avriel,M., Stone,J., Entriken,R., Nakayama,M. (1989):
Decomposition Techniques for Multiarea Generation and Transmission Planning under Uncertainty, EPRI report 2940-1

Davis,P.,J. and Rabinowitz,P. (1984):
Methods of Numerical Integration, Academic Press, London.

Geoffrion,A.M. (1974):
Elements of Large-Scale Mathematical Programming, Management Science 16, No 11.

Glynn,P.W. and Iglehart,D.L. (1988):
Importance Sampling for Stochastic Simulation, submitted for Publication.

Hammersly,J.M. and Handscomb (1964):
Monte Carlo Methods, Mathuen, London.

Kall,P.(1979):
Computational Methods for Two Stage Stochastic Linear Programming Problems, Z. angew. Math. Phys. 30, 261-271

Kall,P. ,Ruszczynski,A., Frauendorfer,K. (1988):
Approximation Techniques in Stochastic Programming, in: Ermoliev,Y. and R.J. Wets (Eds.): Numerical Techniques for Stochastic Optimization

Nazareth,L. and Wets.R.J.:
Algorithms for Stochastic Programs: The Case of Nonstochastic Tenders, Mathematical Programming Study 28, 48-62.

Rockafellar,R.T. and Wets,R.J.(1989):
Scenario and Policy Aggregation in Optimization under Uncertainty, Mathematics of Operations Research (to appear)

Van Slyke and Wets,R.J. (1969):
L-Shaped Linear Programs with Applications to Optimal Control and Stochastic Programming, SIAM Journal of Applied Mathematics, Vol 17, 638-663.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

| REPORT DOCUMENTATION PAGE | | READ INSTRUCTIONS BEFORE COMPLETING FORM |
|---|---|---|
| 1. REPORT NUMBER<br>SOL 89-13 | 2. GOVT ACCESSION NO. | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE (and Subtitle)<br>Monte Carlo (Importance) Sampling within a Benders Decomposition Algorithm for Stochastic Linear Programs | | 5. TYPE OF REPORT & PERIOD COVERED<br>Technical Report<br>6. PERFORMING ORG. REPORT NUMBER |
| 7. AUTHOR(s)<br>Gerd Infanger | | 8. CONTRACT OR GRANT NUMBER(s)<br>N00014-89-J-1659 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS<br>Department of Operations Research - SOL<br>Stanford University<br>Stanford, CA 94305-4022 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS<br>1111MA |
| 11. CONTROLLING OFFICE NAME AND ADDRESS<br>Office of Naval Research - Dept. of the Navy<br>800 N. Quincy Street<br>Arlington, VA 22217 | | 12. REPORT DATE<br>September 1989<br>13. NUMBER OF PAGES<br>30 pages |
| | | 15. SECURITY CLASS. (of this report)<br>UNCLASSIFIED<br>15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |

16. DISTRIBUTION STATEMENT (of this Report)

**This document has been approved for public release and sale; its distribution is unlimited.**

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Linear Programming; Mathematical Programming; Stochastic Programming; Two-Stage Stochastic Linear Programming; Benders' Decomposition; Monte Carlo Sampling; Importance Sampling.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

(see reverse side)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

- ABSTRACT -

**SOL 89-13: Monte Carlo (Importance) Sampling within a Benders' Decomposition Algorithm for Stochastic Linear Programs, Gerd Infanger (September 1989, 30 pp.).**

A method employing decomposition techniques and Monte Carlo sampling (importance sampling) to solve stochastic linear programs is described and applied to capacity expansion planning problems of electric utilities. We consider uncertain availability of generators and transmission lines and uncertain demand. Numerical results are presented.